

## Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of  
Business

Lee Kong Chian School of Business

3-2018

# The predictive power of people's intraindividual variability across situations: Implementing whole trait theory in assessment

Filip LIEVENS

Singapore Management University, [filiplievens@smu.edu.sg](mailto:filiplievens@smu.edu.sg)

Jonas W. B. LANG

Ghent University

Filip DE FRUYT

Ghent University

Myrjam VAN DE VIJVER

Ghent University

Ronald Joachim BLEADOW

Singapore Management University, [RBLEADOW@smu.edu.sg](mailto:RBLEADOW@smu.edu.sg)

**DOI:** <https://doi.org/10.1037/apl0000280>

Follow this and additional works at: [https://ink.library.smu.edu.sg/lkcsb\\_research](https://ink.library.smu.edu.sg/lkcsb_research)

Part of the [Human Resources Management Commons](#), and the [Organizational Behavior and Theory Commons](#)

### Citation

LIEVENS, Filip; LANG, Jonas W. B.; DE FRUYT, Filip; VAN DE VIJVER, Myrjam; and BLEADOW, Ronald Joachim. The predictive power of people's intraindividual variability across situations: Implementing whole trait theory in assessment. (2018). *Journal of Applied Psychology*. 103, (7), 753-771. Research Collection Lee Kong Chian School Of Business.

**Available at:** [https://ink.library.smu.edu.sg/lkcsb\\_research/5846](https://ink.library.smu.edu.sg/lkcsb_research/5846)

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [libIR@smu.edu.sg](mailto:libIR@smu.edu.sg).





# The Predictive Power of People's Intraindividual Variability Across Situations: Implementing Whole Trait Theory in Assessment

Filip Lievens  
Singapore Management University

Jonas W. B. Lang, Filip De Fruyt, Jan Corstjens,  
and Myrjam Van de Vijver  
Ghent University

Ronald Bledow  
Singapore Management University

In the last decade, there has been increased recognition that traits refer not only to between-person differences but also to meaningful within-person variability across situations (i.e., whole trait theory). So far, this broader more contemporary trait conceptualization has made few inroads into assessment practices. Therefore, this study focuses on the assessment and predictive power of people's intraindividual variability across situations. In three studies (either in student or employee samples), both test-takers' mean trait scores and the variability of their responses across multiple written job-related situations of a situational judgment test (SJT) were assessed. Results revealed that people's intraindividual variability (a) was related to their self-rated functional flexibility, (b) predicted performance above their mean scores, and (c) predicted their actual personality state variability over 10 days. These results open opportunities for complementing traditional selection procedures with more dynamic indices in assessment.

*Keywords:* intraindividual variability, personality states, situational judgment test, personal initiative, whole trait theory

In the traditional definition of traits, the notions of “stability” (i.e., referring to continuity across time) and “consistency” (i.e., accounting for the communalities in characteristic behavior observable across situations) constitute pivotal features (Allport, 1927). This trait concept contrasts to a situationist view in which the stability and consistency of behavior across situations are questioned. For decades, these two clashing views have sparked the person-situation debate (Epstein & O'Brien, 1985; Funder & Colvin, 1991; Mischel, 1968).

In more recent times, however, these two perspectives are no longer seen as irreconcilable. In fact, already more than 10 years

ago, Fleeson (2004) posited that the person-situation debate in the personality and social psychology field is over because both sides seem to be correct (see also Fleeson & Nofle, 2009). That is, there is now relative consensus that traits are most useful for predicting behavior over a longer time span, whereas situations make individual behavior vary in the short run. Therefore, Fleeson argued that we “need to embrace a new understanding of traits, realizing that people tend to demonstrate significant flexibility in their behavior [across situations] and that traits are best used for predicting trends” (p. 86).

In line with this suggestion, personality and social psychology research has considerably elaborated its thinking about consistency of trait manifestations across situations and expanded trait theory with contemporary conceptualizations (e.g., personality states) that are based on the axiom that traits refer not only to between-person mean differences in behavioral tendencies but also to meaningful within-person variability across situations. More generally, Fleeson and Jayawickreme (2015) referred to such a trait conceptualization as “whole trait theory” because *both* consistency and within-person variability across situations are considered as meaningful substantive sources of trait information. The overarching idea is that while traits provide a useful summary of a person's general behavioral tendencies (e.g., his or her level of sociability) across many situations, additional information can be obtained if we know how much variability in trait expressions the person displays across various situations (Huang, Ford, & Ryan, 2016).

If we build on a broader and more contemporary trait concept and thus recognize that traits are useful for the description and prediction of differences in behavior as well as that there is

---

Filip Lievens, Lee Kong Chian School of Business, Singapore Management University; Jonas W. B. Lang, Department of Personnel Management and Work and Organizational Psychology, Ghent University; Filip De Fruyt, Department of Developmental, Personality and Social Psychology, Ghent University; Jan Corstjens and Myrjam Van de Vijver, Department of Personnel Management and Work and Organizational Psychology, Ghent University; Ronald Bledow, Lee Kong Chian School of Business, Singapore Management University.

Portions of this article were presented at SIOP 2014 and SIOP 2015. Study 1 and Study 2 were supported by a grant of the Fund for Scientific Research-Flanders (G092512N). We thank Brian Connelly, Eric Heggestad, Cornelius König, Stefan Krumm, and In-Sue Oh for their insightful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Filip Lievens, Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899. E-mail: filiplievens@smu.edu.sg

substantial intraindividual variability in behavior across situations, it follows that we should measure both between and within-person trait variability to improve prediction (Judge, Simon, Hurst, & Kelley, 2014). At present, this more integrative and contemporary trait conceptualization has already made inroads in the organizational behavior domain in which diary studies linked specific episodes and demands to intraindividual variation in people's traits as captured by their personality states (e.g., Bledow, Schmitt, Frese, & Kühnel, 2011; Huang & Ryan, 2011; Judge et al., 2014; Minbashian, Wood, & Beckmann, 2010).

However, the implications of this paradigm shift that emphasizes both the level and the variability of traits have not found their way in the field of assessment, selection, and prediction in I/O psychology. One key reason for why so far this new trait perspective has not been adopted is that in an assessment context it is practically difficult to ask candidates to complete a diary on different moments or to come back on several occasions for repeated measurement. We thus do not know whether an assessment of people's within-person variability across situations is also a potentially useful substantive source of information for improving our predictions over and above traditional trait level scores. Therefore, this study seeks to answer this pressing question and focuses on the added value of an assessment of people's intraindividual variability across situations.

Our objectives are threefold. First, we lay out novel approaches for capturing (via a situational judgment inventory/situational judgment test [SJT]) and quantifying (via Item Response Theory [IRT]) people's within-person variability in an assessment context. Second, we shed light on the conceptual meaning of within-person variability by examining whether an assessment of within-person variability represents not only error variance but also substantive variance. As our third related objective, we investigate whether predictions can be improved with an assessment of people's within-person variability across situations over and above their traditional mean level scores.

## Conceptual Meaning of Intraindividual Variability Across Situations

### Historical Roots and Contemporary Notions

Interest in within-person variability is not new. Researchers have long been interested in the variability of people's behavior (Flugel, 1929; Hollingsworth, 1925; Kehr, 1916) and as early as the 1920s it became apparent that noteworthy individual differences might exist in people's variability (for early reviews, see Spearman, 1927, 1950; for a later review, see Eysenck, 1970). Spearman (1927) and Walton (1936) were arguably some of the first researchers to systematically study individual differences in test-takers' fluctuation in responding to tests and tasks using factor analytic methods. A general "oscillation" (fluctuation) factor emerged in both investigations, which was only weakly correlated with *g*/performance and negatively correlated with a steadiness of character rating provided by teachers in Walton's study. These findings led Spearman (1927, 1950) to postulate individual differences in variability/consistency as a possible additional major factor (aside from the *g* factor).

In the assessment center field, there also exists longstanding interest in assessing people's consistency and variability across the

various assessment center exercises (see Gibbons & Rupp, 2009, for a review). This is best evidenced by the large-scale Management Progress Study that was conducted at AT&T over 25 years. When candidate ratings were factor analyzed, two variability-related factors emerged. One factor was labeled "behavioral flexibility," whereas the other factor was termed "stability of performance" (Bray, Campbell, & Grant, 1974). Both factors were related to extrinsic career success variables such as promotion and salary progress.

Since 2000, there has been a resurgence of interest in within-person variability. This is attested by a contemporary conceptualization of personality that expanded traditional definitions beyond purely dispositional models. In particular, it was acknowledged that the expression of traits can vary across occasions, as reflected in the notion of *personality states*, that is, short-term fluctuations in personality that can be observed as situation-specific expression of traits (Fleeson, 2001, 2007; Fleeson & Gallagher, 2009). In this conceptualization, the trait concept refers not only to enduring mean level differences in behavioral tendencies between individuals, but also captures the stable variability in behavior within the person across situations. Importantly, in various studies, Fleeson and colleagues found that there was at least as much variability within a person in personality states across situations than there was variability between persons. More generally, Fleeson's whole-trait theory integrates personality states within the traditional trait concept. That is, traits are described as distributions of personality states (i.e., as "individuals' accumulation of everyday personality states," Fleeson, 2007, p. 825).

### Different Interpretations

What does within-person variability across situations represent? Generally, two main perspectives on people's variability in scores across situations can be distinguished. Fournier, Moskowitz, and Zuroff (2008) summarized these perspectives as follows: "If error is randomly distributed across occasions, situations . . . , then the process of aggregating or smoothing across assessments should provide a closer approximation of the true score. . . . However, perhaps scores do not need to be aggregated to a single point to capture consistencies in the behavior of individuals. There could be local consistencies in the behavior of individuals that would be smoothed over by aggregating to the mean score. To the extent that an individual's behavior varies systematically and meaningfully across situations, the process of aggregation throws away true variance with error variance."

In other words, relying on the principle of aggregation (Epstein, 1979; Rushton, Brainerd, & Pressley, 1983), one interpretation posits that individuals' variability across situations represents random fluctuation that should be averaged across these situations. Thus, it is then assumed that people's average behavioral tendencies as captured by their mean trait score will lead to the best prediction and that person-situation variability will not add extra predictive power. Recall that the interpretation of within-person variability representing error variance is based on the traditional trait concept that considers traits to encompass enduring dimensions of individual differences in tendencies to show consistent patterns of thoughts, feelings, and actions (Eysenck & Eysenck, 1985; McCrae & Costa, 1995).

Conversely, another interpretation flows from a more contemporary trait conceptualization: Just like dispositions are stable and distinct features of individuals (as indicated by individuals' mean level on constructs), intraindividual variability across situations is also systematic, distinct, and meaningful (Dalal et al., 2015; Fleeson, 2001; Fournier et al., 2008). In other words, it is posited that people's variability across situations does not only represent random fluctuation that should be averaged across those situations but also represents in itself meaningful substantive variance. So, this contemporary trait perspective considers individuals' variability across situations to be also systematic and meaningful variance that should enhance prediction above and beyond their mean scores.

Contrasting these traditional trait and contemporary perspectives leads to our general hypothesis that we will test across three studies: We expect within-person variability in trait expression across situations to represent not only error variance but also substantive meaningful variability and thus to enhance the prediction of performance over and above mean scores. So, our key conceptual contribution is that we focus not only on mean level scores on a trait but also on variability around the trait level, which is expected to provide a more comprehensive assessment and improve prediction.

## Measurement of Intraindividual Variability Across Situations

### How to Capture Intraindividual Variability?

In experience-sampling research, a target person's self-report ratings across different days typically serve as basis for capturing within-person variability (e.g., Fleeson, 2001). Given the practical difficulties of embedding a diary approach in assessment practices, Dalal et al. (2015) called for the development of new assessment methods for capturing within-person variability. Essentially, these methods should have the following characteristics: (a) presenting a variety of situations to people, (b) capturing people's reactions to these situations, (c) linking their reactions to underlying traits, and (d) computing a variability score on the traits to assess the variation in reactions across situations (apart from a mean score on the trait across situations).

More specifically, Dalal et al. (2015) suggested that adaptations of assessment center (ACs, aka high fidelity simulations) or SJTs (aka low fidelity simulations) might be well suited to operationalize these principles. For example, the SJT method presents a multitude of situations to test-takers and prompts for their reactions to them. Moreover, an advantage over experience sampling is that SJT situations are standardized and can be chosen to represent relevant job situations (Lievens, in press). Echoing these recommendations, Huang and Ryan (2011) posited that "new tools may be developed that more directly assess situational contingencies, such as situational judgment inventories focused on assessing specific traits that also vary situational characteristics systematically" (p. 480).

Consistent with these recommendations, our studies present test-takers with a set of written situations followed by response options that vary in terms of trait levels (higher vs. lower). In line with the broader more contemporary trait concept, this allows

computing people's mean level scores as well as their variation across situations and using these mean level *as well as* variability scores for prediction purposes.

### How to Measure Intraindividual Variability?

Recently, Dalal et al. (2015) grouped the variety of approaches for measuring people's within-person variability into three broad categories. The first category encompassed statistical operationalizations. For example, a standard deviation is typically computed in experience sampling (diary) studies of personality states (e.g., Fleeson & Gallagher, 2009). Apart from such statistical operationalizations, content-general operationalizations are also used for measuring intraindividual variability across situations. These content-general operationalizations refer to the use of individual differences measures that capture the extent to which one's behavior is influenced by situational cues. Examples of such content-general operationalizations include self-monitoring (Gangestad & Snyder, 1985) or functional flexibility (Paulhus & Martin, 1988). The third category was labeled content-independent operationalizations and included measures such as attitude strength (e.g., Pomerantz, Chaiken, & Tordesillas, 1995).

Of these three approaches, Dalal et al. (2015) noted that statistical operationalizations such as the standard deviation (*SD*) of a person's scores across situations are the most straightforward (one does not need to administer an additional measure) and direct (unobtrusive) operationalizations of within-person variability (see also Cole, Bedeian, Hirschfeld, & Vogel, 2011; Roberson, Sturman, & Simons, 2007). That is also the reason why *SD* scores have been commonly used in experience-sampling (e.g., Fleeson & Gallagher, 2009). However, some researchers have criticized the *SD* as a variability measure (e.g., Cucina & Vasilopoulos, 2005). In particular, when an SJT is used for capturing intraindividual variability, the *SD* might also encompass other sources of variance. It might not only reflect the degree of variability with which a person shows a tendency to endorse/choose/express responses related to a given trait across situations. Possibly, the *SD* score is also affected by the difficulty of the situations (items) and the trait estimate of the person (this follows from the idea that measurement error fluctuates across the trait continuum in IRT, cf. Hambleton, Swaminathan, & Rogers, 1991).<sup>1</sup>

To address this issue, we introduce an IRT approach on the basis of IRT tree models as a method to model variability across situations. Specifically, we built on recent work on multiple process IRT or tree models. An IRT tree model (Böckenholt, 2012; De Boeck & Partchev, 2012; Tutz, 1990) is a multidimensional IRT

<sup>1</sup> Given that experience-sampling studies use the same scale each day and these scales have a larger number of items/observations, the ratings are less likely affected by the situation/item difficulty problem. However, a disadvantage of experience-sampling operationalizations is that the situations included in the measurement for each person vary/fluctuate so the validity of the *SD* as a measurement of behavior depends on a number of assumptions that a standardized situational inventory does not require. Most notably, in an experience-sampling operationalization, one needs to assume that (a) the situations sampled are representative of all possible situations for each person and (b) that the variability in the situations is a result of the person's behavior instead of the person's environment (Fleeson & Law, 2015). A person that works in an office space with many other people, for instance, is more likely to be confronted with a greater variety of interpersonal situations than a person working alone at home.



model. Tree models were originally developed in the context of choice modeling economics (Hensher, Rose, & Greene, 2015) and feature prominently in the work of Nobel Prize winner Daniel McFadden (McFadden, 2001). Tree models split the information from an item into several pseudoitems that represent distinct behavioral tendencies. The approach is particularly interesting for research on variability because IRT tree models for ratings recently described in the literature (Böckenholt, 2012; Zettler, Lang, Hülshager, & Hilbig, 2016) include pseudoitems that functionally separate variability from the standing on the trait. Our approach for operationalizing within-person variability draws upon this approach and splits the information from situational ratings into (a) *persons'* mean-trait level/latent trait, (b) *persons'* tendency to show variability on the trait, (c) *items'* mean-level/latent difficulty, and (d) *items'* tendency to elicit variability in ratings. As a key implication, this approach permits conceptual clarity because it disentangles people's intraindividual variability from other sources of variance (as opposed to the *SD*).

In short, use of an SJT is an approach that can complement and extend earlier approaches for capturing within-person variability. At the same time, this approach has also unique challenges for quantifying within-person variability that we dealt with via our IRT approach.

## Study 1

Study 1 aims to shed light on the conceptual meaning of within-person variability by examining whether a statistically derived within-person variability score correlates with self-reported variability ratings. In addition, we examine whether such a within-person variability score predicts performance over and above mean scores.

To examine these issues, we develop an SJT for potential use in student admissions testing. In the last decade, there has been an increasing interest in educational admissions to complement cognitive measures with SJTs for broadening skill coverage and increasing diversity (e.g., Lievens, Buyse, & Sackett, 2005; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). The SJT developed presents a variety of situations to test-takers together with response options that represent different trait levels of two traits: sociability and dutifulness. These two traits were chosen for their relevance in academic settings (Oswald et al., 2004; Poropat, 2009). For example, Oswald et al.'s model of academic performance distinguishes between two broad dimensions: intrapersonal and interpersonal performance. To predict the intrapersonal component, we selected dutifulness given its consistent validity in academic and training contexts. We opted for sociability to predict the interpersonal component (e.g., work in small-group class projects).

## Conceptual Meaning of Intraindividual Variability

As noted above, one aim of Study 1 consists of shedding light on the conceptual meaning of within-person variability by examining whether intraindividual variability scores relate to self-ratings about the extent to which one's behavior is variable and influenced by situational cues (aka content-general operationalizations of intraindividual variability, Dalal et al., 2015). Evidence for such a relationship then indicates that within-person variability represents also substantive and not only error variance.

Paulhus and Martin (1988) conducted a conceptual and empirical comparison of various individual differences measures of people's intraindividual variability across situations. Among the measures examined, self-monitoring and functional flexibility seem especially relevant for this study's purpose. Snyder (1974) characterized a self-monitor as someone who "out of a concern for social appropriateness, is particularly sensitive to the expression and self-presentation of others in social situations and uses these cues as guidelines for monitoring his own self-presentation" (p. 528). Due to their ability and motivation to observe and regulate their behavior and self-presentation to social cues, high self-monitors typically show less consistency and thus more intraindividual variability in their responses across different social situations than low self-monitors whose behavior is more guided by their inner state. Given this link between self-monitoring and the tendency to shift one's response to the specific social context, we propose:

*Hypothesis 1a (H1a):* Intraindividual variability will be significantly related to self-monitoring ratings.

Besides self-monitoring, Paulhus and Martin (1988) also highlighted the importance of functional flexibility. They defined this construct as "the ability to adjust one's behavior to the interpersonal demands of a wide range of situations" (p. 91, see also Erickson, Newman, & Pincus, 2009). At a conceptual level, functional flexibility captures self-perceptions of situational appropriateness as well as self-perceptions of a large response repertoire for adjusting to different interpersonal demands. Paulhus and Martin's (1987) measure of functional flexibility asks people to self-report how capable they are of showing various behaviors when the situation requires it. People also have to assess the difficulty and anxiety of performing these behaviors, and the tendency to avoid situations demanding such behavior. We expect that people who rate themselves high on functional flexibility, and thus report having a wide repertoire of available behaviors that they can appropriately adjust to the situational demands, will show higher intraindividual variability in their responses to situations. Thus:

*H1b:* Intraindividual variability will be significantly related to functional flexibility ratings.

## Added Value of Intraindividual Variability in Predicting Performance

Another aim of Study 1 is to examine the predictive potential of an assessment of within-person variability. We expect that within-person variability will be important for both traits of interest in this study: sociability and dutifulness. Regarding sociability, we base our expectation on interpersonal theory: In everyday interactions at work, it is important to not be constrained to certain types of interpersonal behaviors and to show some degree of flexibility along interpersonal dimensions (Kiesler, 1983). Some interpersonal situations require a controlled and restrained way of interacting, whereas others call for warm and open interpersonal interactions. Similarly, sometimes employees need to take the lead in task-related interactions; at other times they need to follow the instructions of others (Tiedens & Fragale, 2003). More generally, according to interpersonal theory, responding to a diversity of interpersonal situations and dealing with different interaction part-

ners requires considerable behavioral flexibility (Kiesler, 1983; Tiedens, Unzueta, & Young, 2007; Wiggins, 1991).

Regarding dutifulness, our expectation of intraindividual variability being beneficial is grounded in allocation of resources theory (Beal, Weiss, Barros, & MacDermid, 2005; Minbashian & Luppino, 2014). This theory predicts that people are most effective when they can allocate all of their resources to the task at hand, when this is demanded by the task. Many tasks require dutifulness; however, when a task does not demand dutifulness (e.g., a task contributes little to organizational effectiveness), it may be more adaptive to conserve resources and express lower dutifulness. In support of this, Minbashian et al. (2010) found that people whose state conscientiousness was more contingent on task demands performed better because they were able to increase their conscientious-related behavior in the face of more urgent and more difficult tasks (see also Huang & Bramble, 2016). In academic settings too, students might be seen as more effective when they are more dutiful and diligent when the situation's urgency and difficulty call for it. In short, we expect that people who display variability in responses related to the traits of interest in this study will receive higher performance ratings. In addition, we expect that such variability will show incremental validity over their mean scores.

We also expect that people's intraindividual variability will explain not only additional variance over their mean scores but will also show incremental validity over their self-monitoring and functional flexibility ratings in predicting performance. This expectation is based on the consideration that the use of a statistical operationalization of within-person variability is less prone to (un)intentional response distortion than self-reports of self-monitoring and functional flexibility. In addition, use of a statistically derived measure of intraindividual variability depends less on people's ability to introspect. Even if people do not intentionally distort their responses, they simply might not have a good idea of their self-monitoring and functional flexibility and use different anchors when they compare themselves with others. Importantly, if this expectation of incremental variance is supported, the use of a statistically derived score of within-person variability might serve as a new angle for measuring adaptability. Thus:

*H2a:* Intraindividual variability will be significantly related to peer-rated performance.

*H2b:* Intraindividual variability will have incremental validity for predicting performance over mean scores and self-reported variability (self-monitoring and functional flexibility) ratings.

## Method

**Sample and procedure.** Two-hundred and thirty five third-year psychology students from a large West-European university were invited to participate in this study for course credit (31% men, average age = 21.41 years). They were also promised to receive a detailed feedback report. All of these students had taken Industrial and Organizational Psychology (I/O) as their major. They completed among others<sup>2</sup> a demographic questionnaire, a personality inventory, and the SJT. To measure their motivation, they also filled out the test motivation scale (see Arvey, Strickland, Drauden, & Martin, 1990) after the SJT. This scale consisted of three items (e.g., "Doing well on this test was important to me")

with a Likert-type scale from 1 (*very inaccurate*) to 5 (*very accurate*). The internal consistency reliability of the ratings was .74 and the mean score was 4.07 ( $SD = .62$ ), underscoring students' motivation to do well on the SJT. Afterward, we asked them to distribute rating forms to two classmates with whom they had worked in team projects. Data were collected in accordance with the ethical guidelines for research involving human subjects by Ghent University, Belgium (G092512N).

**Development of SJT.** The development of the SJT consisted of three main steps. First, we gathered a variety of relevant situations related to sociability and dutifulness (Tett & Burnett, 2003). To this end, we utilized two sources of information. As one source of information, we scrutinized other measures that were developed for educational contexts (e.g., Motowidlo, Hooper, & Jackson, 2006; Oswald et al., 2004; Peeters & Lievens, 2005). As another source of information, six subject matter experts (graduate psychology students; 67% female; average age = 22.5 years) generated critical incidents of student performance. Next, we converted the situations gathered into 24-item stems.

In the second stage, we wrote four response options per item stem. Response options reflected higher versus lower levels of the traits (Motowidlo & Beier, 2010; Motowidlo et al., 2006). To guide this process, each response option was written as an operationalization of a trait adjective (e.g., helpful, curt, accommodating, suspicious). To determine whether the adjective reflected higher/lower levels of the trait, we relied on a well-validated adjective classification system (i.e., Abridged Big Five Circumplex [AB5C]; De Raad, Hendriks, & Hofstee, 1992; Hofstee, De Raad, & Goldberg, 1992; Woods & Anderson, 2016). Table 1 presents an example item and the associated adjectives.

In the third step, seven graduate I/O psychology students (five females, two males) participated in a retranslation procedure to verify whether each response option reflected its respective adjective. When the correct retranslation's percentage was lower than 60%, we revised the options. Twenty-six of the 96 response options needed revision. A second retranslation procedure with a similar group of graduate I/O psychology students confirmed that our revisions were successful.

This development process resulted in a 24-item SJT, with each item having four response options. To examine alternate-form reliability we added three item clones (constructed via an item-cloning procedure, see Lievens & Sackett, 2007). So, the final SJT consisted of 27 items. A behavioral tendency response instruction ("What would you do?") was used and the response format was a Likert-type scale, ranging from 1 (*not likely at all*) to 5 (*very likely*). All SJT items were in Dutch (i.e., the participants' native language).

**Level and variability scores on SJT.** As noted above, a challenge for researchers that seek to measure intraindividual variability in the context of an SJT is to appropriately separate various sources of variance (i.e., person's trait level, item's trait level, person variance, item variance). We addressed this issue by fitting a tree model, which we refer to as the trait variability tree

<sup>2</sup> A cognitive ability test was also included. Intraindividual variability and cognitive ability correlated  $-.13$  ( $p < .05$ ). This result should be interpreted with caution because the cognitive ability test scores were range restricted in this population of third-year students. Inclusion of cognitive ability in the analyses did not change the results.



Table 1  
Example of SJT Item

Item stem
You are a member of a team of four students. In the context of an assignment, each of the students has to conduct an interview with a practitioner prior to a predetermined date. Afterwards, each team has to integrate the interviews into common findings. One of your fellow students who has a reputation of always being too late tells you that she has already conducted her interview. However, you doubt that this is true. What would you do?
a. As the interview is only part of the assignment, you propose helping her with transcribing her interview (higher level of sociability; "helpful").
b. You tell her that you do not believe her and you repeat that the deadline applies to everyone, including her (lower level of sociability; "curt").
c. This is not worth a quarrel. So, you let this go (higher level of sociability; "accommodating").
d. You ask her to prove that the interview took place because you do not trust her (lower level of sociability; "suspicious").

Note. SJT = situational judgment test. The text between parentheses was not shown to test-takers. All SJT items were in Dutch (i.e., the participants' native language). So, this example item was translated.

model (TVTM). This model is a modified version of Böckenholt's three-pseudoitem model (Böckenholt, 2012). Recently, this model has been introduced to the personality literature for analyzing Likert-scale type data (Zettler et al., 2016): It suggests that people typically answer Likert-scales by first deciding on whether the midcategory applies (i.e., whether they should give a directed answer at all; Pseudoitem I); when they decide not to apply the midcategory, they next decide on which direction their answer will be (Pseudoitem II); finally they decide on whether they endorse the "agree" or the "strongly agree" category (Pseudoitem III). In this type of model, the level of the trait is captured functionally independent from the variability on the trait. Variability is captured both by a person's tendency to prefer a directional response over the neutral response and a person's tendency to prefer the extreme "strongly agree" and "strongly disagree" categories over the "agree" and "disagree" categories. For the trait variability tree model (TVTM), we constrained the two variability pseudoitems (Pseudoitem I and Pseudoitem III) such that both tendencies were captured by one underlying trait. This constraint of the model is justified because the two variability processes were highly correlated in earlier personality data (Zettler et al., 2016) and in our data when we fitted the model without the constraint (.59). Furthermore, the two pseudoitems are also conceptually similar because both decisions capture variability in trait responses that is functionally independent from trait level. We also found intraindividual variability to be highly consistent across the two traits (.98). The size of this correlation among intraindividual variability scores of the two traits is not surprising because this is a robust finding in personality (e.g., Baird et al., 2006; Eid & Diener, 1999; Fleeson, 2001) and in earlier research (Spearman, 1927, 1950). Consistent with the high correlation among variability in sociability and dutifulness and with prior research, we therefore use an overall intraindividual variability model with an overall variability score in our analyses.

In sum, the IRT model we used for Study 1 was a model with three individual differences dimensions: A sociability trait, a dutifulness trait, and overall variability. The full model specification as well as model estimates are provided in Appendix A. After fitting the IRT model, we estimated empirical Bayes estimates using the IRT model and the glmer function in the lme4 package for R (Bates, Mächler, Bolker, & Walker, 2015). These scores were used in our analyses.

We also conducted two sets of analyses to assess classic reliability. First, we computed the internal consistencies of the ratings. Internal consistencies were .62 and .52 for sociability and dutiful-

ness scores, respectively. These values are in the same range as the mean internal consistency values of .46 and .57 for SJT scores in the meta-analyses of Catano, Brochu, and Lamerson (2012) and Campion, Ployhart, and MacKenzie (2014). Second, we examined alternate-form reliability using the scores on the three original items and their respective cloned items (see above). Alternate form reliabilities were .65 and .76 for sociability and dutifulness scores, respectively, which is satisfactory given that these values are based on a small set of items.

#### Other measures.

**Traditional personality inventory.** We included a self-report personality inventory that measured three personality traits (agreeableness, extraversion, and conscientiousness). These traits were chosen because they were conceptually related to either sociability (agreeableness, extraversion; Leising & Bleidorn, 2011; McCrae & Costa, 1989) or dutifulness (conscientiousness). We took 10 items for each of the three traits from the International Personality Item Pool (Goldberg, 1999). These items were rated on a scale ranging from 1 (*very inaccurate*) to 7 (*very accurate*). Internal consistencies of these scales' ratings were .83 (A), .89 (C), and .90 (E).

**Self-monitoring.** We used the revised 18-item scale of Gangestad and Snyder (1985) to measure self-monitoring. A correct/incorrect response format was used. An example item was "I would probably make a good actor." The internal consistency reliability for the scale's ratings was .70.

**Functional flexibility.** Paulhus and Martin (1987) developed the Battery of Interpersonal Capabilities (BIC) to assess people's self-perceptions in variability across interpersonal situations. The BIC is an 80-item measure based on 16 behavioral tendencies. Per behavioral tendency, people report how likely they could display the behavior when the situation required it, how difficult they find this, how anxious they feel when performing it, and their tendency to avoid situations that elicit the behavior. An example item was: "How capable are you of being dominant when the situation requires it?" Items were rated on a 7-point Likert scale, from 1 (*not at all*) to 7 (*very much*). The internal consistency of the scale's ratings was .79. As could be expected, self-monitoring and functional flexibility were significantly related to each other (.23, see Table 2).

**Criterion measure.** Peer-rated criteria are appropriate when validating interpersonal measures in academic contexts because peers typically use an interpersonal lens in their assessment of others' task accomplishment (Oswald et al., 2004). We collected ratings of overall performance from two peers. The ratings' eco-

Table 2

*Descriptive Statistics and Correlations of Study 1 and 2 Variables*

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Personality inventory																
1. Extraversion (E)	5.02	1.00														
2. Agreeableness (A)	5.87	.64	.31**													
3. Conscientiousness (C)	4.80	1.12	.06	.26**												
SJT																
4. Mean Sociability	-.00	.15	.27**	.31**	-.06											
5. Mean Dutifulness	.01	.22	.04	.13*	.28**	.18**										
6. Intraindividual variability	-.00	.37	.17**	.25**	.23**	.06	-.26**									
Other measures																
7. Functional flexibility	5.38	.49	.30**	.10	.09	.05	-.03	.27**								
8. Self-monitoring	9.22	3.20	.41**	-.07	-.12	.01	-.15*	.09	.23**							
9. Overall performance	3.92	.51	.16*	.16*	.39**	-.04	.20**	.19*	.08	-.03						
10-day Diary <sup>a</sup>																
10. State E Mean	3.10	.21	.35**	.13	.11	.20*	.15	.05	-.04	-.02	.14					
11. State A Mean	3.54	.27	-.06	.27**	.28**	.19*	.10	.18*	-.10	-.25**	.10	.16				
12. State C Mean	3.56	.27	-.02	.20*	.27**	.17	.20*	.06	-.01	-.14	.30**	.24**	.53**			
13. State E <i>SD</i>	.38	.14	-.02	.11	-.11	-.03	-.08	.22*	.04	-.03	-.06	-.08	-.23*	-.26**		
14. State A <i>SD</i>	.33	.12	.01	.06	-.12	-.07	-.15	.21*	.03	.15	-.03	-.14	-.40**	-.27**	.36**	
15. State C <i>SD</i>	.45	.16	.01	.05	-.18*	-.04	-.18	.21*	.02	-.10	-.12	-.13	-.24**	-.38**	.45**	.45**

Note. SJT = situational judgment test. *N* = 232, except 10-day diary study *n* = 120; The personality inventory and functional flexibility were rated on a 7-point scale, whereas the SJT and the overall performance measure were rated on a 5-point scale. Self-monitoring was measured on a dichotomous scale.  
<sup>a</sup> Based on 1,200 measurement occasions (10 for each person).

\*  $p < .05$ . \*\*  $p < .01$ .

logical validity was ensured because the students rated peers with whom they had engaged in team-based school assignments (*Mdn* = 19 months of acquaintance). Overall performance was measured via Williams and Anderson's (1991) measure, which consists of seven items with a 5-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is "Performs his/her duties thoroughly and strives for perfection." The internal consistency of the ratings was .84, which is similar to the mean intrarater reliability of peer ratings (.85) in the meta-analysis of Viswesvaran, Ones, and Schmidt (1996). The interrater reliability of the two peers' ratings was .49 (cf. .42 in Viswesvaran et al., 1996). Our analyses used the average of the peer ratings.

## Results

**Mean scores: Construct-related and criterion-related validity.** Given that we developed a new SJT for measuring sociability and dutifulness, it was important to start with inspecting the correlations between the mean SJT scores and conceptually related self-report personality ratings. As shown in Table 2, all convergent validity correlations were significant. Sociability correlated .27 and .31 with extraversion and agreeableness, respectively, whereas dutifulness correlated .28 with conscientiousness. The size of these correlations conforms to what one might expect given that the SJT presents specific situations to activate responses related to sociability and dutifulness, whereas in the personality inventory people self-report on personality traits in a generic fashion (Back & Egloff, 2009). These convergent validity correlations were also higher than the correlations between the mean SJT scores and conceptually unrelated personality ratings (–.06, .04, and .13), attesting to the discriminant validity of these mean SJT scores.

In terms of criterion-related validity, conscientiousness (.39) emerged as best predictor among the self-report personality rat-

ings. Self-reported ratings of extraversion and agreeableness performed worse (both .16). A similar pattern was observed for the SJT in which the mean dutifulness score correlated .20 with peer-rated performance. The mean sociability score was not a significant predictor.

In short, both the construct-related and criterion-related validities of mean SJT scores match meta-analytic findings of the link between SJTs and personality (Christian, Edwards, & Bradley, 2010; McDaniel, Hartman, Whetzel, & Grubb, 2007). So, we continued with testing our hypotheses about the meaning and added value of an assessment of intraindividual variability via the SJT.

**Intraindividual variability: Test of hypotheses.** To test hypotheses related to intraindividual variability, the possibility of statistical dependence between mean and variability scores should first be examined (Cole et al., 2011). When mean and *SD* scores are used, researchers have argued that the variability scores cannot be interpreted because it is possible that standing on the trait and variability are systematically confounded. As noted previously, our IRT approach addresses this approach by disentangling these aspects. In any case, in this study, the correlations between the two mean scores and the intraindividual variability score were small to moderate (.06 and –.26).<sup>3</sup>

The first set of hypotheses proposed that intraindividual variability would significantly correlate with self-report ratings of self-monitoring (H1a) and functional flexibility (H1b). Table 2 shows that the correlation between intraindividual variability and functional flexibility was significant (.27) but the one (.09) with

<sup>3</sup> We also ran a set of supplementary analyses in which we restricted the correlations between variability and the two trait scores to be orthogonal. The conclusions of the study did not change. Therefore, we report the unrestricted results.

self-monitoring was not. None of the correlations between the mean SJT scores and functional flexibility/self-monitoring was significant. These results support H1b that an intraindividual variability score captures substantive variance related to functional flexibility. There was no support for H1a about the relation between intraindividual variability and self-monitoring.

H2a posited that intraindividual variability would be significantly correlated with performance. There was support for H2a because intraindividual variability was significantly correlated with performance (.19). H2b dealt with the incremental validity of intraindividual variability over and above mean scores and functional flexibility/self-monitoring. We conducted a hierarchical regression in which we entered the mean sociability and dutifulness scores, the flexibility/self-monitoring ratings, and finally the intraindividual variability score.

As shown in Table 3, intraindividual variability accounted for incremental variance ( $R^2$  increment of .06) in predicting performance, which lends support to H2b. Additional analyses supported the robustness of these findings.<sup>4</sup> For example, a hierarchical regression with conceptually related mean self-report personality scores (instead of the mean SJT scores) entered as the first block confirmed the  $R^2$  increment of the intraindividual variability score reported above.

## Discussion

One objective of Study 1 was to test the hypothesis that intraindividual variability would correlate with self-report measures of variability. Results generally supported this hypothesis, lending credence to the notion of within-person variability representing not only error variance but also substantive variance. The relationship of intraindividual variability with functional flexibility was stronger than the one with self-monitoring. This result might be explained by the difference between self-monitoring and functional flexibility. That is, self-monitoring deals primarily with individuals' self-presentation motives instead of with the breadth of their behavioral repertoire (Paulhus & Martin, 1988). In addition, self-monitoring zooms into the tendency to adjust behavior to situations in general but does not specify how this might be accomplished in terms of a varied range of specific behaviors (which is captured via the functional flexibility measure).

As another objective, Study 1 investigated whether an assessment of within-person variability added predictive validity to traditional mean scores. As a key finding, there was evidence for substantial additional predictive power when within-person variability was added. It is insightful to link these validity results of intraindividual variability to the predictive validity record of personality traits. For example, the validity record of agreeableness and extraversion which are the Big Five traits related to sociability and interpersonal skills (Leising & Bleidorn, 2011; McCrae & Costa, 1989) is mixed. Our results suggest that one explanation behind the mixed findings for the validity of interpersonally related personality traits might be that it makes sense to be sociable in some situations but not in other situations. This explanation is also in line with interpersonal theory (Kiesler, 1983; Wiggins, 1991).

A limitation of Study 1 is that we did not relate intraindividual variability as measured via an SJT to people's actual variability in real-life. Therefore, Study 2 investigates whether assessing peo-

ple's within-person variability (as captured by the variability in their scores on an SJT) is predictive of their actual within-person variability in personality states.

## Study 2

Study 2 links people's intraindividual variability on the SJT to actual variation in their personality states across days. As noted by Fleeson (2001), personality states reflect short-term fluctuations in personality that result from situation-specific expressions of traits. To gather personality states, people are typically asked to complete a diary in which they characterize their personality state on several occasions during a day (e.g., Fleeson, 2001; Fleeson & Gallagher, 2009) or daily (e.g., Judge et al., 2014).

It is important to note that personality state variability should be conceptually distinguished from within-person variability as captured via our SJT. This intraindividual variability is based on people's choices of response options to written situations and deals with variability in responding across situations due to a trait being differentially activated, suppressed, or expressed based on characteristics of the situation. Conversely, personality states relate to changes (increases or decreases) in action tendencies across situations (Fleeson & Jayawickreme, 2015). The measurement is also different. One approach is based on the use of a low-fidelity simulation, whereas a diary method captures self-reports of actual behavior. Note too that we gathered the personality state data 20 months after the SJT. Given these differences, an examination of a link between people's intraindividual variability in choosing responses (across written situations) and their variability in personality states (across real situations) is a strong test of the validity of an assessment of intraindividual variability via an SJT. So, we formulate a research question instead of a hypothesis:

**Research question.** Will intraindividual variability across written situations of an SJT predict variation in personality states in everyday life?

## Method

**Sample.** Study 2 consisted of a subsample of Study 1. About 20 months later, all subjects from the original study were contacted again via email and invited to participate in a follow-up daily diary study in exchange for a gift card of choice (€30,00). When participants consented, they received an email at the end of each weekday (around 5 p.m.) with a link to the online diary measure (see below). Subjects could fill out the survey on their computer or handheld device at any point during the evening up until 2 a.m. They received such daily e-mails until they had completed the survey for 10 weekdays.

One-hundred and thirty five participants agreed to participate (57% response rate). As some participants did not complete the measures for all 10 weekdays or had to be dropped due to technical problems with registering their responses, we had usable responses of 120 participants (23% men, average age = 23.2 years, 76%

<sup>4</sup> Entering quadratic and cubic terms prior to the variability score to test for curvilinear effects did not change the results. Note further that adding the interaction between mean and intraindividual variability scores was neither significant for sociability,  $t(228) = 1.50$ ;  $p = .14$  nor for dutifulness,  $t(228) = 1.74$ ;  $p = .08$ .

Table 3  
Summary of Hierarchical Regression Analyses of Study 1 for Predicting  
Peer-Rated Performance

Predictor	<i>b</i>	<i>beta</i>	Beta 95% CI [LL, UL]	<i>r</i>	Fit	Difference
1. Intercept	3.80**					
Mean sociability	-.36	-.11	[-.23, .02]	-.04		
Mean dutifulness	.65**	.28	[.15, .42]	.20**		
Functional flexibility	.02	.02	[-.11, .15]	.08		
Self-monitoring	-.00	-.01	[-.14, .11]	-.03	$R^2 = .054^*$	
2. Intraindividual variability	.36**	.27	[.13, .40]	.19**	$R^2 = .114^{**}$	$\Delta R^2 = .06^{**}$

Note.  $N = 232$ ; *b* represents unstandardized regression weights; *beta* indicates the standardized regression weights; *r* represents the zero-order correlation. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

\*  $p < .05$ . \*\*  $p < .01$ .

full-time students, 24% working either part-time or full-time). Drop out analyses revealed no significant differences between the diary study group and the original sample on self-reported personality (see Table 2) or on other variables, with one exception: The diary sample included significantly more females.

**Personality states via diary study.** We developed a diary study according to the experience-sampling methodology (Beal, 2015; Fisher & To, 2012). For the duration of 10 weekdays, participants had to describe themselves once during the evening on 24 adjectives. Similar to other personality state assessment studies (e.g., Fleeson & Gallagher, 2009), we used adjective-based Big Five measures on which participants described their behavior during the day (instead of describing themselves in general). Hereby we relied on the same 24 adjectives that we used in the development of the SJT (see Study 1). To avoid response sets, the presentation order of the adjectives differed each day. Per adjective, participants were asked to what extent the adjective described themselves during the day on a 5-point Likert scale, ranging from 1 (*completely not characteristic*) to 5 (*very characteristic*). Mean and *SD* scores for each of the three personality states were computed per participant per day. Across the 10 measurement occasions, internal consistency reliability estimates associated with these scores ranged from .66 to .85.

## Results

Our research question dealt with whether people's intraindividual variability in responding to the situations of the SJT predicted their actual variability in personality states. Given that these personality states were measured via a diary approach, we followed recommendations in the experience-sampling literature (Nezlek, 2012, see also Footnote 1) and therefore used people's within-person standard *SD* across days as a measure of state variability.<sup>5</sup> All regression analyses used the mean ratings across days as control. For completeness, we also always added functional flexibility and self-monitoring as controls.

Table 4 shows that intraindividual variability in people's responses on the SJT (determined via the same IRT approach as in Study 1) predicted variation in similar personality states in everyday life. The intraindividual variability score had incremental validity over and above functional flexibility, self-monitoring, and

mean scores in predicting variability in personality states across days ( $R^2$  increment from .04 to .07).<sup>6</sup>

## Discussion

Study 2 provided further evidence for the meaning and importance of an assessment of people's intraindividual variability across situations. We found evidence that the overall intraindividual variability picked up *actual* real-life variability in personality states, even though the intraindividual variability score was based on within-person variability to *written* situations (aka a low-fidelity simulation). To our knowledge, this is the first study that links results on an SJT to dynamic real-life criteria nearly two years later.

It should be noted, though, that the results of Study 1 and Study 2 are based on student samples that completed an SJT that we specifically designed for testing our hypotheses. This SJT consists of situation descriptions followed by response options that operationalize different levels of the two traits (sociability and dutifulness). As the SJT resembles construct-driven SJTs (Christian et al., 2010), one might wonder whether an assessment of people's intraindividual variability also provides additional predictive information in *existing* construct-driven SJTs. Therefore, we conducted Study 3 in an employee sample in which we tested our general hypothesis by reanalyzing the data of Bledow and Frese's (2009) SJT. Such an investigation is important because it might

<sup>5</sup> When the variable is normally distributed, the *SD* becomes increasingly normally distributed as the number of observations per *SD* increases (Kenney & Keeping, 1951; Roesslein, Wolf, Wampfler, & Wegscheider, 2007). In this study, plots of the residuals' distribution showed little deviation from normality. Note that multilevel analyses that modeled residual variance patterns (Hoffman, 2007) yielded the same results.

<sup>6</sup> We ran the same analyses with respondents that were not fully employed. For this subsample, results were similar:  $\Delta R^2 = .08^{**}$   $\Delta R^2 = .05^*$ , and  $\Delta R^2 = .05^*$  for predicting the extraversion *SD*, agreeableness *SD*, and conscientiousness *SD*, respectively. In another analysis, we used an aggregate of the three diary-based *SD* measures as an overall dependent variable. In Step 1, we entered the state extraversion, agreeableness, and conscientiousness means as predictors in addition to mean sociability, mean dutifulness, functional flexibility, and self-monitoring. In Step 2, we entered the intraindividual variability SJT index. In Step 1,  $R^2$  was .21 ( $p < .01$ ). In Step 2,  $\Delta R^2$  was .10 ( $p < .01$ ). So, both of these extra analyses confirmed our results.



Table 4

Summary of Hierarchical Regression Analyses of Study 2 for Predicting Deviations in Diary Reports Across 10 Days

Predictor	<i>b</i>	<i>beta</i>	Beta 95% CI [LL, UL]	<i>r</i>	Fit	Difference
Variability in extraversion in diary reports						
1. Intercept	.54*					
State extraversion Mean	-.05	-.08	[-.27, .11]	-.08		
Mean sociability	-.04	-.05	[-.24, .14]	-.03		
Mean dutifulness	-.01	-.01	[-.21, .18]	-.08		
Functional flexibility	.01	.02	[-.17, .21]	.04		
Self-monitoring	-.00	-.05	[-.24, .14]	-.03	$R^2 = .017$	
2. Intraindividual variability	.09*	.23	[.04, .42]	.22*	$R^2 = .062$	$\Delta R^2 = .05^*$
Variability in agreeableness in diary reports						
1. Intercept	1.12**					
State agreeableness Mean	-.20**	-.45	[-.62, -.27]	-.40**		
Mean sociability	-.01	-.01	[-.18, .16]	-.06		
Mean dutifulness	-.01	-.02	[-.20, .15]	-.15		
Functional flexibility	-.02	-.07	[-.24, .10]	.03		
Self-monitoring	.00	.05	[-.12, .22]	.15	$R^2 = .178^{**}$	
2. Intraindividual variability	.10**	.29	[.12, .47]	.21*	$R^2 = .251^{**}$	$\Delta R^2 = .07^{**}$
Variability in conscientiousness in diary reports						
1. Intercept	1.36**					
State conscientiousness mean	-.25**	-.40	[-.57, -.23]	-.38**		
Mean sociability	-.01	-.01	[-.18, .16]	-.04		
Mean dutifulness	-.05	-.06	[-.24, .12]	-.18		
Functional flexibility	.01	.03	[-.14, .21]	.02		
Self-monitoring	-.01*	-.18	[-.35, -.00]	-.10	$R^2 = .184^{**}$	
2. Intraindividual variability	.10*	.22	[.04, .39]	.21*	$R^2 = .225^{**}$	$\Delta R^2 = .04^*$

Note.  $N = 120$ ; *b* represents unstandardized regression weights; *beta* indicates the standardized regression weights; *r* represents the zero-order correlation. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

\*  $p < .05$ . \*\*  $p < .01$ .

show that our results are not limited to our specific SJT but generalize to other SJTs (as long as those SJTs present response options that operationalize different trait levels).

### Study 3

Bledow and Frese (2009) developed an SJT of personal initiative (SJT-PI) to measure people's tendency to behave proactively in work settings. The concept of personal initiative was introduced by Frese and Fay (2001) and refers to self-starting actions people initiate to bring about desired individual and organizational outcomes. Personal initiative is a concrete behavioral expression of the trait called proactive personality (Crant, 1995; Spitzmuller, Sin, Howe, & Fatimah, 2015). To capture this trait with an SJT, the SJT-PI presents a set of scenarios with four to five response options that vary in personal initiative and asks respondents to select the responses they would most likely display. Response options high in personal initiative concern self-starting actions to deal with job-related problems and are contrasted with less proactive response options.

We examine here if the tendency to display variability in personal initiative predicts job performance. More specifically, variability in choosing response options high or low in personal initiative across work scenarios of the SJT-PI may add to the prediction of job performance over and above a person's mean tendency. Past research on proactivity suggests that a mean tendency to display personal initiative is related to job performance (e.g., Bledow & Frese, 2009; Seibert, Kraimer, & Crant, 2001).

When respondents display a preference for personal initiative across a wide range of job situations, they will tend to take initiative when a situation at work provides the opportunity and thereby contribute to personal and organizational effectiveness.

However, personal initiative may not always be the most effective response in organizations and a one-sided focus on self-starting actions can be dysfunctional for job performance. High job performance often requires following others' instructions and adapting to their initiatives (Grant, Gino, & Hofmann, 2011; Griffin, Neal, & Parker, 2007). Interpersonal theory suggests that flexibility with respect to proactive and reactive actions is necessary for interpersonal effectiveness (Kiesler, 1983; Tiedens et al., 2007). For instance, employees often need to react and help others instead of proactively pursuing goals. Moreover, employees can encounter problems at work for which taking personal initiative is ineffective, for instance when they do not have the abilities to solve a problem or when taking initiative comes at the cost of neglecting other tasks (Chan, 2006). For such situations, acceptance of the status quo and responses such as emotion regulation may be more effective than proactive attempts to initiate change (Gross & John, 2003).

In sum, this line of reasoning suggests that intraindividual variability in personal initiative may contribute to job performance in addition to the mean tendency to display personal initiative. Intraindividual variability in personal initiative implies that a person displays personal initiative selectively and is not focused exclusively on proactive attempts to reach desired goals.

*H3a:* Intraindividual variability in personal initiative will be positively related to supervisor ratings of job performance.

*H3b:* Intraindividual variability in personal initiative will have incremental validity for predicting supervisor ratings of job performance over mean scores in personal initiative.

## Method

**Sample and procedure.** The sample consisted of 126 employees in the financial industry who completed 14 items of the SJT-PI for research purposes. For each item, participants indicated which response they would most likely display. Response options varied in whether they reflected low (−1), moderate (0) or high (+1) personal initiative (see Bledow & Frese, 2009 for details about item development).<sup>7</sup> Data were collected in accordance with the ethical guidelines for research involving human subjects by University of Freiburg, Germany.

We adapted the TVTM IRT model from Study 1 and Study 2 to this format. As shown in the Appendix A, this model is a reduced form of the model used in Study 1 and Study 2 because there is only one trait and three Likert response options. Again, we first estimated the IRT model and then estimated empirical Bayes (also known as factor or theta) scores for both variability and latent trait level for each person. The trait score captures respondents' mean tendency to select response options with higher personal initiative, while the variability score captures the functionally distinct tendency to select a high or low personal initiative response option rather than a response option with moderate personal initiative. In the IRT model (see also the estimates in Appendix A), the variability and trait level scores were moderately correlated (.37, see Appendix B). We therefore fitted an alternative IRT model in which the two scores were not allowed to correlate. The results were similar and thus we report below the results for the unrestricted model just like in Study 1 and Study 2.

**Criterion measures.** To assess overall job performance we used a composite of the mean of two criterion dimensions: task performance and contextual performance. Task performance was measured with three items of a behavioral-anchored rating scale developed by Motowidlo and Van Scotter (1994). An example item asked managers to assess on a 7-point scale whether an employee's performance fell below, met or exceeded performance standards ( $\alpha = .96$ ). Contextual performance (i.e., helping) was measured on a 5-point scale with five items developed by Organ and Konovsky (1989) and adapted by Van Dyne and LePine (1998). An example item was 'This employee is always willing to help and support others' ( $\alpha = .89$ ). People's SJT ratings and their supervisor assessments were collected concurrently. Supervisor ratings of job performance were available for 77 employees.

## Results

Table 5 presents the descriptive statistics of this study. The relationship between mean personal initiative and supervisor ratings of job performance was .28 ( $p < .001$ ), whereas the relationship between intraindividual variability in personal initiative and job performance was .37 ( $p < .001$ ). Table 6 presents hierarchical regression analyses with job performance and its two facets task performance and contextual performance as outcomes. The out-

comes were first regressed on mean personal initiative. In a second model, intraindividual variability in personal initiative was added. In support of Hypothesis 3, variability in personal initiative predicted job performance ratings by managers after controlling for mean personal initiative ( $\Delta R^2 = .08$ ). We next ran separate analyses for task performance and organizational behavior. For task performance, intraindividual variability in personal initiative was positively related to the outcome, however, it did not explain incremental variance after controlling for people's mean tendency. By contrast, intraindividual variability in personal initiative added to the prediction of contextual performance ( $\Delta R^2 = .10$ ). In sum, intraindividual variability in personal initiative was thus a predictor of job performance, because it helped to explain whether a person engaged in contextual performance.

## Discussion

Study 3 further confirms our general hypothesis with an existing SJT and supervisor ratings of job performance. It shows that including an assessment of within-person variability (in this case in personal initiative) adds to the prediction of job performance. While we found no relationship of variability in personal initiative with task performance, the relationship with contextual performance, namely helping was significant. In line with interpersonal theory, variability in a person's tendency to display personal initiative was important in the interpersonal domain. Presumably, flexibility in a person's tendency to be proactive is important for helping, because helping can occur proactively but also as a reaction to the request of others. For performing individual tasks, by contrast, a mean tendency to display personal initiative appeared to be sufficient.

## General Discussion

### Implications for Theory

The traditional model underlying assessment and prediction is based on a trait conceptualization that stresses consistency and stability across situations, thereby considering variability of no utility. However, there is increased recognition in various fields that dynamic trait constructs (i.e., people's intraindividual variability across situations) might complement this traditional conceptualization and improve understanding and prediction of relevant criteria (e.g., Côté, Moskowitz, & Zuroff, 2012; Fleeson & Law, 2015; Judge et al., 2014; Minbashian & Luppino, 2014).

Our results speak in favor of such a more comprehensive and integrative approach to assessment because we found evidence for the validity of both mean level and variability scores for predicting performance. Importantly, there was also evidence that the intraindividual variability score picked up variability in personality states in real-life, even though this score was based on within-person variability in responses to written situations. At a conceptual level, these results support assumptions underlying more contemporary trait conceptualizations that emphasize that within-

<sup>7</sup> Bledow and Frese (2009) originally also used a second parallel rating for each item in which participants indicated their least likely response. Given the complex psychometric dependency between the two ratings, we focused only on the most likely option in the IRT analyses.



Table 5  
Descriptive Statistics and Correlations of Study 3 Variables

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. Mean personal initiative	−.02	.60				
2. Intraindividual variability	−.00	.08	.68**			
3. Contextual performance	3.88	.67	.23*	.38**		
4. Task performance	4.82	1.26	.27*	.29**	.63**	
5. Overall job performance	−.00	1.00	.28*	.37**	.90**	.90**

Note. *N* = 77.

\* *p* < .05. \*\* *p* < .01.

person variability across situations should be included (apart from the traditional cross-situational consistency). So, the essence is that in a more comprehensive assessment two main components are of central importance, namely the average trait-related response across situations and the variable expression of trait-related responses to different situations (how people adapt their responses to the demands of the situation). This expansion adds a more dynamic assessment perspective to the traditional static approaches, which has also utility in resulting in more predictive power.

As another implication, this study provides insights as to what intraindividual variability means and when intraindividual variability is adaptive because we were able to disentangle reasons behind it. In fact, in the personality state literature, it has been difficult to unravel the competing explanations for intraindividual variability across situations (Erickson et al., 2009; Fleeson & Gallagher, 2009; Fleeson & Law, 2015; Wrzus & Mehl, 2015). For example, does high variability in extraversion reflect (a) exposure to or selection of varied situations calling for different levels of extraverted behaviors (in these situations anyone would show this much variability in extraversion in similarly varied situations); (b) sensitivity to contextual cues and ability to appropriately adapt behavior accordingly (see Mischel & Shoda, 1995); or (c) capricious and unsystematic behavior across situations (Baird, Le, & Lucas, 2006; Erickson et al., 2009)? Interestingly, the standardized format of SJT items allows ruling out explanation (a). As we

correlated SJT scores with a performance outcome, we were also able to disentangle the other explanations. That is, our results show that intraindividual variability was positively and not negatively related to performance (explanation b instead of explanation c). In our study, the adaptive nature of intraindividual variability for interpersonal skills can be understood on the basis of interpersonal theory (Kiesler, 1983; Wiggins, 1991), whereas the allocation of resources theory is a useful framework for understanding why levels of dutifulness should be matched to task demands (Beal et al., 2005; Minbashian & Luppino, 2014).

On a more general level, the issue of the (mal)adaptive nature of intraindividual variability has a long history (e.g., in clinical and social psychology, Bleidorn & Ködding, 2013; Erickson et al., 2009; in assessment centers, see the behavioral flexibility” vs. “stability of performance” factors identified in the Management Progress Study; Brayet et al., 1974). We posit that a lot depends on the nature of the construct of interest. On one hand, take a construct such as safety-rule adherence. For this construct, maximum adherence seems the only acceptable form of behavior given the possible consequences for performance. When intraindividual variability for such constructs is purely stimulus driven, or driven by impulses/mood states, it likely has negative consequences. In some studies, negative effects of intraindividual variability also occurred in the form of an interaction effect, where smaller intraindividual variability reflects more consistency. For example, in Fleisher, Woehr, Edwards, and Cullen (2011), conscientiousness was more positively associated with performance when variability in conscientiousness was lower.

On the other hand, when variability results from a person’s integrated past experiences and the tacit knowledge of different behaviors being effective in different circumstances, then variability is likely to be adaptive. This article’s constructs (dutifulness, sociability, and personal initiative) likely belong to this adaptive category because these constructs were measured with items focusing on work environments and included behaviors that depend on people’s resources such that adaptive regulation was necessary and beneficial for performance.

Table 6  
Summary of Hierarchical Regression Analyses of Study 3 for Predicting Task, Contextual, and Overall Job Performance

Predictor	<i>b</i>	<i>beta</i>	Beta 95% CI [LL, UL]	<i>r</i>	Fit	Difference
Task performance						
1. Intercept	4.83**					
Mean personal initiative	.55*	.25	[.03, .47]	.26*	<i>R</i> <sup>2</sup> = .067*	
2. Intraindividual variability	3.53	.17	[−.05, .39]	.19	<i>R</i> <sup>2</sup> = .097*	Δ <i>R</i> <sup>2</sup> = .03
Contextual performance						
1. Intercept	3.88**					
Mean personal initiative	.21	.18	[−.04, .39]	.20	<i>R</i> <sup>2</sup> = .042	
2. Intraindividual variability	3.52**	.32	[.11, .54]	.34**	<i>R</i> <sup>2</sup> = .146**	Δ <i>R</i> <sup>2</sup> = .10**
Overall job performance						
1. Intercept	.00					
Mean personal initiative	.42*	.24	[.02, .45]	.26*	<i>R</i> <sup>2</sup> = .066*	
2. Intraindividual variability	4.48*	.28	[.06, .49]	.29**	<i>R</i> <sup>2</sup> = .141**	Δ <i>R</i> <sup>2</sup> = .08*

Note. *N* = 77; *b* represents unstandardized regression weights; *beta* indicates the standardized regression weights; LL and UL indicate the lower and upper limits of a confidence interval, respectively; *r* represents the zero-order correlation.

\* *p* < .05. \*\* *p* < .01.

## Implications for Future Research

Our results open a window of opportunity for future research. First, we encourage researchers to use other approaches for testing this study's basic premise that within-person variability across situations represents not only noise but also substance. An SJT constitutes only one of the possible formats. Specifically, we welcome future studies to test this study's key assumptions about the importance of intraindividual variability with instruments that measure actual behavior such as assessment center exercises. For example, one might develop short 3-min assessment center exercises (Brannick, 2008) that activate specific dimensions to study candidates' variability across them. This would also enable linking our findings to the extensive body of research on candidates' cross-situational inconsistency in assessment centers (see, e.g., Gibbons & Rupp, 2009; Lance, 2008; Lievens, 2002; Putka & Hoffman, 2013).

Second, now that we established that within-person variability across written situations represents not only error variance but also substantive and meaningful variance, future research should examine other dynamic indices (Moskowitz & Zuroff, 2004). Specifically, if we cast our study in Cronbach and Gleser's (1953) seminal conceptualization of *elevation* (average tendency across situations), *scatter* (standard deviation across situations), and *shape* (patterns of variations across situations), our study dealt with the first two aspects. So, the next important step consists of uncovering patterns in people's within-person variability. Situation-trait contingencies denote systematic relationships between a particular situation feature and a trait (Huang & Ryan, 2011) exemplify such patterns. So, for instance, one could examine under which specific situations an individual is more variable and whether this pattern fits the job profile (Gibbons & Rupp, 2009; Lievens, in press).

Meaningful intraindividual patterns of situation-behavior relationships are also known as behavioral signatures or as "if-then" situation-behavior profiles (Fournier et al., 2008; Mischel & Shoda, 1995; Shoda, Mischel, & Wright, 1994). The notion of behavioral signatures acknowledges that an individual might choose different response options depending on his or her signature and distinctive way of construing the specific situational features. To examine intraindividual patterns of situation-trait contingencies, it is best to rely on a situational taxonomy (e.g., Parrigon, Woo, Tay, & Wang, 2017; Rauthmann et al., 2014) with specific situational features as a basis for the development of the situational descriptions. In addition, to allow examination of situation-trait contingencies, situational inventories with a sufficient number of items should be constructed.

In this study, the situations of our SJT in Study 1 were derived from subject matter experts and prior SJTs. Accordingly, in Study 1 we are not able to relate people's responses to specific situational features. In Study 3, it is possible to sort the SJT items for personal initiative in three groups (i.e., organizational, individual, and ambiguous). Therefore, we report here exploratory analyses to examine potential situation-trait contingencies. Specifically, we tested the idea that intraindividual variability would vary between these three item types by specifying a version of our IRT model in which the latent trait for variability interacts with the situation characteristics such that the model estimates a separate latent trait for each variability. This model did not provide a better fit to the data than

a model without this interaction,  $\chi^2(df = 7) = 2.05$ ,  $p = .96$ . However, note that in these analyses the sample size in both items and persons is limited.

Third, future studies should examine whether the assessment of people's within-person variability can also predict other criterion components. In this study, self-reported actual personality state variability across days and performance ratings provided by peers and supervisors served as external variables. In the future, we recommend linking intraindividual variability to adaptive and leadership performance (Huang, Ryan, Zabel, & Palmer, 2014; Minbashian et al., 2010; Pulakos, Arad, Donovan, & Plamondon, 2000).

## Limitations

Apart from the limitations mentioned per study, some general limitations should also be acknowledged. First, our studies were conducted in a low-stakes context with either third-year psychology students or employees. Although the test motivation of the psychology students was relatively high, confirmation of our results in high-stakes contexts and other samples is needed. Second, our studies dealt with only three traits (sociability, dutifulness, and personal initiative) and two situational inventories. So, our hypotheses should be tested with other traits and other situational measures. Third, all of our results are based on a statistical IRT-based operationalization of within-person variability. Future studies are needed to test our hypotheses with other operationalizations of intraindividual variability (Dalal et al., 2015).

## Implications for Practice

Given that this study presents the first investigation of the added value of an assessment of within-person variability, we warn beforehand that the following practical implications should be cautiously interpreted. As a first implication, this study translated recent more comprehensive trait conceptualizations into a concrete assessment practice. That is, we developed a straightforward and practically feasible approach for capturing intraindividual variability on the basis of responses to an SJT because it presents a multitude of situation descriptions with the level of detail needed for studying intraindividual fluctuations (Lievens, in press). This novel approach also presents practitioners with additional meaningful information for making selection decisions. Accordingly, the use of a statistically derived variability score provides a new angle (beyond traditional self-reports that might be prone to lack of self-insight and/or impression management) for assessing the much-needed 21st century skill of adaptability. This extra information is also beneficial for test-takers because it can be added to their feedback reports.

A requirement of the SJT is that the response component (i.e., response options) is designed so that the response options carefully represent construct-related courses of action to a broad set of job-related situations. This requirement should not be a deal breaker for practitioners. In fact, it echoes repeated calls to use a more construct-driven approach in designing SJTs (Labrador, Christiansen, & Burns, 2006; Lievens & Motowidlo, 2015; Motowidlo & Beier, 2010; Motowidlo et al., 2006; Stemler & Sternberg, 2006). The fact that an SJT with construct-related response options allows assessing variability might serve as an additional argument for developing such SJTs.

Finally, this study presented a new approach for scoring SJTs. Specifically, we showed how IRT can be used to derive a mean level and a variability score from SJTs. A major advantage of our approach is that the mean level and variability score are functionally independent because each rating is split into several pseudoi-tems. Furthermore, the approach controls for item difficulty. It should be noted that the new approach does not require substantial technical knowledge or specialized software because the models described can be fitted via the freely available R language (the authors can be contacted for the code) and various software packages such as SAS or Mplus (see Böckenholt, 2012, for examples).<sup>8</sup>

## Conclusion

This study began integrating a broader and more contemporary trait conceptualization into assessment and prediction in applied settings. As key results, we found that people's intraindividual variability across different written situations (a) was measureable via an IRT-based approach, (b) reflected substantive variance in the form of self-reported functional flexibility and actual personality state variability across days, and (c) incrementally added to the prediction of traditionally used mean level scores. As far as we know, this is the first study that uncovers the conceptual and predictive importance of such an assessment of intraindividual variability. We encourage future researchers and practitioners to take these findings as point of departure toward complementing traditional selection practices in organizations with more dynamic measures.

<sup>8</sup> Practitioners that still want to use the traditional *SD* for quantifying intraindividual variability should realize that *SD* and IRT approaches show more convergence when mean and variability scores are (a) not highly correlated and (b) computed on the basis of a large number of data points. In that case, the difficulty differences among the items/item options average out and have less profound effect on *SD*.

## References

- Allport, G. W. (1927). Concepts of trait and personality. *Psychological Bulletin*, 24, 284–293. <http://dx.doi.org/10.1037/h0073629>
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43, 695–716. <http://dx.doi.org/10.1111/j.1744-6570.1990.tb00679.x>
- Back, M. D., & Egloff, B. (2009). Yes we can! A plea for direct behavioral observation in personality research. *European Journal of Personality*, 23, 403–405.
- Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*, 90, 512–527. <http://dx.doi.org/10.1037/0022-3514.90.3.512>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, 2, 383–407. <http://dx.doi.org/10.1146/annurev-orgpsych-032414-111335>
- Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology*, 90, 1054–1068. <http://dx.doi.org/10.1037/0021-9010.90.6.1054>
- Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology*, 62, 229–258. <http://dx.doi.org/10.1111/j.1744-6570.2009.01137.x>
- Bledow, R., Schmitt, A., Frese, M., & Kühnel, J. (2011). The affective shift model of work engagement. *Journal of Applied Psychology*, 96, 1246–1257. <http://dx.doi.org/10.1037/a0024532>
- Bleidorn, W., & Ködding, C. (2013). The divided self and psychological (mal)adjustment—A meta-analytic review. *Journal of Research in Personality*, 47, 547–552. <http://dx.doi.org/10.1016/j.jrp.2013.04.009>
- Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, 17, 665–678. <http://dx.doi.org/10.1037/a0028111>
- Brannick, M. T. (2008). Back to basics of test construction and scoring. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 131–133. <http://dx.doi.org/10.1111/j.1754-9434.2007.00025.x>
- Bray, D. W., Campbell, R. J., & Grant, D. L. (1974). *Formative years in business: A long-term AT&T study of managerial lives*. New York, NY: Wiley.
- Campion, M. C., Ployhart, R. E., & MacKenzie, W., Jr. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance*, 27, 283–310. <http://dx.doi.org/10.1080/08959285.2014.929693>
- Catano, V. M., Brochu, A., & Lamerson, C. D. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, 20, 333–346. <http://dx.doi.org/10.1111/j.1468-2389.2012.00604.x>
- Chan, D. (2006). Interactive effects of situational judgment effectiveness and proactive personality on work perceptions and work outcomes. *Journal of Applied Psychology*, 91, 475–481. <http://dx.doi.org/10.1037/0021-9010.91.2.475>
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83–117. <http://dx.doi.org/10.1111/j.1744-6570.2009.01163.x>
- Cole, M. S., Bedeian, A. G., Hirschfeld, R. R., & Vogel, B. (2011). Dispersion-composition models in multilevel research: A data-analytic framework. *Organizational Research Methods*, 14, 718–734. <http://dx.doi.org/10.1177/1094428110389078>
- Côté, S., Moskowitz, D. S., & Zuroff, D. C. (2012). Social relationships and intraindividual variability in interpersonal behavior: Correlates of interpersonal spin. *Journal of Personality and Social Psychology*, 102, 646–659. <http://dx.doi.org/10.1037/a0025313>
- Crant, J. M. (1995). The proactive personality scale and objective job performance among real estate agents. *Journal of Applied Psychology*, 80, 532–537. <http://dx.doi.org/10.1037/0021-9010.80.4.532>
- Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, 50, 456–473. <http://dx.doi.org/10.1037/h0057173>
- Cucina, J. M., & Vasilopoulos, N. L. (2005). Nonlinear personality-performance relationships and the spurious moderating effects of trait-edness. *Journal of Personality*, 73, 227–259. <http://dx.doi.org/10.1111/j.1467-6494.2004.00309.x>
- Dalal, R. S., Meyer, R. D., Bradshaw, R. P., Green, J. P., Kelly, E. D., & Zhu, M. (2015). Personality strength and situational influences on behavior: A conceptual review and research agenda. *Journal of Management*, 41, 261–287. <http://dx.doi.org/10.1177/0149206314557524>
- De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73, 533–559. <http://dx.doi.org/10.1007/s11336-008-9092-x>
- De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software*, 48, 1–28. <http://dx.doi.org/10.18637/jss.v048.c01>



- De Raad, B., Hendriks, A. A. J., & Hofstee, W. K. B. (1992). Towards a refined structure of personality traits. *European Journal of Personality*, 6, 301–319. <http://dx.doi.org/10.1002/per.2410060405>
- Eid, M., & Diener, E. (1999). Intraindividual variability in affect; Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, 76, 662–676. <http://dx.doi.org/10.1037/0022-3514.76.4.662>
- Epstein, S. (1979). Stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097–1126. <http://dx.doi.org/10.1037/0022-3514.37.7.1097>
- Epstein, S., & O'Brien, E. J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin*, 98, 513–537. <http://dx.doi.org/10.1037/0033-2909.98.3.513>
- Erickson, T. M., Newman, M. G., & Pincus, A. L. (2009). Predicting unpredictability: Do measures of interpersonal rigidity/flexibility and distress predict intraindividual variability in social perceptions and behavior? *Journal of Personality and Social Psychology*, 97, 893–912. <http://dx.doi.org/10.1037/a0016515>
- Eysenck, H. J. (1970). *The structure of human personality* (3rd ed.). London, UK: Routledge.
- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science*. New York, NY: Plenum Press. <http://dx.doi.org/10.1007/978-1-4613-2413-3>
- Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior*, 33, 865–877. <http://dx.doi.org/10.1002/job.1803>
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distribution of states. *Journal of Personality and Social Psychology*, 80, 1011–1027. <http://dx.doi.org/10.1037/0022-3514.80.6.1011>
- Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Research*, 13, 83–87. <http://dx.doi.org/10.1111/j.0963-7214.2004.00280.x>
- Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, 75, 825–861. <http://dx.doi.org/10.1111/j.1467-6494.2007.00458.x>
- Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, 97, 1097–1114. <http://dx.doi.org/10.1037/a0016786>
- Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality*, 56, 82–92. <http://dx.doi.org/10.1016/j.jrp.2014.10.009>
- Fleeson, W., & Law, M. K. (2015). Trait enactments as density distributions: The role of actors, situations, and observers in explaining stability and variability. *Journal of Personality and Social Psychology*, 109, 1090–1104. <http://dx.doi.org/10.1037/a0039517>
- Fleeson, W., & Nofle, E. E. (2009). The end of the person-situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, 2, 1667–1684. <http://dx.doi.org/10.1111/j.1751-9004.2008.00122.x>
- Fleisher, M. S., Woehr, D. J., Edwards, B. D., & Cullen, K. L. (2011). Assessing within-person personality variability via frequency estimation: More evidence for a new measurement approach. *Journal of Research in Personality*, 45, 535–548. <http://dx.doi.org/10.1016/j.jrp.2011.06.009>
- Flugel, J. C. (1929). Practice, fatigue, and oscillation. *British Journal of Psychology*, 7, 13.
- Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2008). Integrating dispositions, signatures, and the interpersonal domain. *Journal of Personality and Social Psychology*, 94, 531–545. <http://dx.doi.org/10.1037/0022-3514.94.3.531>
- Frese, M., & Fay, D. (2001). Personal initiative (PI): An active performance concept for work in the 21st century. *Research in Organizational Behavior*, 23, 133–187. [http://dx.doi.org/10.1016/S0191-3085\(01\)23005-6](http://dx.doi.org/10.1016/S0191-3085(01)23005-6)
- Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, 60, 773–794. <http://dx.doi.org/10.1037/0022-3514.60.5.773>
- Gangestad, S., & Snyder, M. (1985). “To carve nature at its joints”: On the existence of discrete classes in personality. *Psychological Review*, 92, 317–349. <http://dx.doi.org/10.1037/0033-295X.92.3.317>
- Gibbons, A. M., & Rupp, D. E. (2009). Dimension consistency as an individual difference: A new (old) perspective on the assessment center construct validity debate. *Journal of Management*, 35, 1154–1180. <http://dx.doi.org/10.1177/0149206308328504>
- Goldberg, L. R. (1999). *International personality item pool: A scientific collaboratory for the development of advanced measures of personality traits and other individual differences*. Retrieved from <http://ipip.ori.org/>
- Grant, A. M., Gino, F., & Hofmann, D. A. (2011). Reversing the extraverted leadership advantage: The role of employee proactivity. *Academy of Management Journal*, 54, 528–550. <http://dx.doi.org/10.5465/AMJ.2011.61968043>
- Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal*, 50, 327–347. <http://dx.doi.org/10.5465/AMJ.2007.24634438>
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85, 348–362. <http://dx.doi.org/10.1037/0022-3514.85.2.348>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Thousand Oaks, CA: Sage.
- Hensher, D., Rose, J., & Greene, W. (2015). *Applied choice analysis*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781316136232>
- Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, 42, 609–629. <http://dx.doi.org/10.1080/00273170701710072>
- Hofstee, W. K. B., de Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63, 146–163. <http://dx.doi.org/10.1037/0022-3514.63.1.146>
- Hollingsworth, H. L. (1925). Correlations of achievement within an individual. *Journal of Experimental Psychology*, 8, 190–208. <http://dx.doi.org/10.1037/h0065414>
- Huang, J. L., & Bramble, R. J. (2016). Trait, state, and task-contingent conscientiousness: Influence on learning and transfer. *Personality and Individual Differences*, 92, 180–185. <http://dx.doi.org/10.1016/j.paid.2015.12.043>
- Huang, J. L., Ford, J. K., & Ryan, A. M. (2016). Ignored no more: Within-person variability enables better understanding of training transfer. *Personnel Psychology*, 70, 557–596. <http://dx.doi.org/10.1111/peps.12155>
- Huang, J. L., & Ryan, A. M. (2011). Beyond personality traits: A study of personality states and situational contingencies in customer service jobs. *Personnel Psychology*, 64, 451–488. <http://dx.doi.org/10.1111/j.1744-6570.2011.01216.x>
- Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: A meta-analytic investigation. *Journal of Applied Psychology*, 99, 162–179. <http://dx.doi.org/10.1037/a0034285>
- Judge, T. A., Simon, L. S., Hurst, C., & Kelley, K. (2014). What I experienced yesterday is who I am today: Relationship of work motivations and behaviors to within-individual variation in the five-factor

- model of personality. *Journal of Applied Psychology*, 99, 199–221. <http://dx.doi.org/10.1037/a0034485>
- Kehr, T. (1916). Versuchsanordnung zur experimentellen Untersuchung einer kontinuierlichen Aufmerksamkeitsleistung [A study setup for the experimental investigation of a continuous attentional performance]. *Zeitschrift für angewandte Psychologie*, 11, 465–479.
- Kenney, J. F., & Keeping, E. S. (1951). *Mathematics of statistics* (2nd ed.). Princeton, NJ: Van Nostrand.
- Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. *Psychological Review*, 90, 185–214. <http://dx.doi.org/10.1037/0033-295X.90.3.185>
- Labrador, J. R., Christiansen, N. D., & Burns, G. N. (2006, May). *Measuring personality using situational judgment tests*. Paper presented at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 84–97. <http://dx.doi.org/10.1111/j.1754-9434.2007.00017.x>
- Leising, D., & Bleidorn, W. (2011). Which are the basic meaning dimensions of observable interpersonal behavior? *Personality and Individual Differences*, 51, 986–990. <http://dx.doi.org/10.1016/j.paid.2011.08.003>
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology*, 87, 675–686. <http://dx.doi.org/10.1037/0021-9010.87.4.675>
- Lievens, F. (in press). Assessing personality–situation Interplay in personnel selection: Toward more integration into personality research. *European Journal of Personality*.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442–452. <http://dx.doi.org/10.1037/0021-9010.90.3.442>
- Lievens, F., & Motowidlo, S. J. (2015). Situational Judgment Tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 3–22. <http://dx.doi.org/10.1017/iop.2015.71>
- Lievens, F., & Sackett, P. R. (2007). Situational judgment tests in high-stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology*, 92, 1043–1055. <http://dx.doi.org/10.1037/0021-9010.92.4.1043>
- McCrae, R. R., & Costa, P. T., Jr. (1989). The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of Personality and Social Psychology*, 56, 586–595. <http://dx.doi.org/10.1037/0022-3514.56.4.586>
- McCrae, R. R., & Costa, P. T., Jr. (1995). Trait explanations in personality psychology. *European Journal of Personality*, 9, 231–252. <http://dx.doi.org/10.1002/per.2410090402>
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91. <http://dx.doi.org/10.1111/j.1744-6570.2007.00065.x>
- McFadden, D. (2001). Economic choices. *The American Economic Review*, 91, 351–378. <http://dx.doi.org/10.1257/aer.91.3.351>
- Minbashian, A., & Luppino, D. (2014). Short-term and long-term within-person variability in performance: An integrative model. *Journal of Applied Psychology*, 99, 898–914. <http://dx.doi.org/10.1037/a0037402>
- Minbashian, A., Wood, R. E., & Beckmann, N. (2010). Task-contingent conscientiousness as a unit of personality at work. *Journal of Applied Psychology*, 95, 793–806. <http://dx.doi.org/10.1037/a0020016>
- Mischel, W. (1968). *Personality and assessment*. London, UK: Wiley.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102, 246–268. <http://dx.doi.org/10.1037/0033-295X.102.2.246>
- Moskowitz, D. S., & Zuroff, D. C. (2004). Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology*, 86, 880–893. <http://dx.doi.org/10.1037/0022-3514.86.6.880>
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, 95, 321–333. <http://dx.doi.org/10.1037/a0017975>
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91, 749–761. <http://dx.doi.org/10.1037/0021-9010.91.4.749>
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475–480. <http://dx.doi.org/10.1037/0021-9010.79.4.475>
- Nezlek, J. B. (2012). *The SAGE library in social and personality psychology methods*. London, UK: Sage. <http://dx.doi.org/10.4135/9781446287903>
- Organ, D. W., & Konovsky, M. (1989). Cognitive versus affective determinants of organizational citizenship behavior. *Journal of Applied Psychology*, 74, 157–164. <http://dx.doi.org/10.1037/0021-9010.74.1.157>
- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, 89, 187–207. <http://dx.doi.org/10.1037/0021-9010.89.2.187>
- Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, 112, 642–681. <http://dx.doi.org/10.1037/pspp0000111>
- Paulhus, D. L., & Martin, C. L. (1987). The structure of personality capabilities. *Journal of Personality and Social Psychology*, 52, 354–365. <http://dx.doi.org/10.1037/0022-3514.52.2.354>
- Paulhus, D. L., & Martin, C. L. (1988). Functional flexibility—A new conception of interpersonal flexibility. *Journal of Personality and Social Psychology*, 55, 88–101. <http://dx.doi.org/10.1037/0022-3514.55.1.88>
- Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, 65, 70–89. <http://dx.doi.org/10.1177/0013164404268672>
- Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. *Journal of Personality and Social Psychology*, 69, 408–419. <http://dx.doi.org/10.1037/0022-3514.69.3.408>
- Poropat, A. E. (2009). A meta-analysis of the Five-Factor model of personality and academic performance. *Psychological Bulletin*, 135, 322–338. <http://dx.doi.org/10.1037/a0014996>
- Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, 85, 612–624. <http://dx.doi.org/10.1037/0021-9010.85.4.612>
- Putka, D. J., & Hoffman, B. J. (2013). Clarifying the contribution of assessee-, dimension-, exercise-, and assessor-related effects to reliable and unreliable variance in assessment center ratings. *Journal of Applied Psychology*, 98, 114–133. <http://dx.doi.org/10.1037/a0030887>
- Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., . . . Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, 107, 677–718. <http://dx.doi.org/10.1037/a0037250>
- Roberson, Q. M., Sturman, M. C., & Simons, T. L. (2007). Does the measure of dispersion matter in multilevel research? A comparison of

- the relative performance of dispersion indexes. *Organizational Research Methods*, 10, 564–588. <http://dx.doi.org/10.1177/1094428106294746>
- Roesslein, M., Wolf, M., Wampfler, B., & Wegscheider, W. (2007). A forgotten fact about the standard deviation. *Accreditation and Quality Assurance*, 12, 495–496. <http://dx.doi.org/10.1007/s00769-007-0285-2>
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity—The principle of aggregation. *Psychological Bulletin*, 94, 18–38. <http://dx.doi.org/10.1037/0033-2909.94.1.18>
- Seibert, S. E., Kraimer, M. L., & Crant, J. M. (2001). What do proactive people do? A longitudinal model linking proactive personality and career success. *Personnel Psychology*, 54, 845–874. <http://dx.doi.org/10.1111/j.1744-6570.2001.tb00234.x>
- Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, 67, 674–687. <http://dx.doi.org/10.1037/0022-3514.67.4.674>
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 526–537. <http://dx.doi.org/10.1037/h0037039>
- Spearman, C. E. (1927). *The abilities of man*. London, UK: Macmillan.
- Spearman, C. E. (1950). *Human ability*. London, UK: Macmillan.
- Spitzmuller, M., Sin, H.-P., Howe, M., & Fatimah, S. (2015). Investigating the uniqueness and usefulness of proactive personality in organizational research: A meta-analytic review. *Human Performance*, 28, 351–379. <http://dx.doi.org/10.1080/08959285.2015.1021041>
- Stemler, S. E., & Sternberg, R. J. (2006). Using Situational Judgment Tests to measure practical intelligence. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 107–131). Mahwah, NJ: Erlbaum.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88, 500–517. <http://dx.doi.org/10.1037/0021-9010.88.3.500>
- Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology*, 84, 558–568. <http://dx.doi.org/10.1037/0022-3514.84.3.558>
- Tiedens, L. Z., Unzueta, M. M., & Young, M. J. (2007). An unconscious desire for hierarchy? The motivated perception of dominance complementarity in task partners. *Journal of Personality and Social Psychology*, 93, 402–414. <http://dx.doi.org/10.1037/0022-3514.93.3.402>
- Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical & Statistical Psychology*, 43, 39–55. <http://dx.doi.org/10.1111/j.2044-8317.1990.tb00925.x>
- Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: Evidence of construct and predictive validity. *Academy of Management Journal*, 41, 108–119. <http://dx.doi.org/10.2307/256902>
- Viswesvaran, C., Ones, D. S., & Schmidt, E. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557–560. <http://dx.doi.org/10.1037/0021-9010.81.5.557>
- Walton, R. D. (1936). Relations between amplitude of oscillations in short period efficiency and steadiness of character. *British Journal of Psychology*, 27, 181–188.
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl: Vol. 2: Personality and psychopathology* (pp. 89–113). Minneapolis, MN: University of Minnesota Press.
- Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17, 601–617. <http://dx.doi.org/10.1177/014920639101700305>
- Woods, S. A., & Anderson, N. R. (2016). Toward a periodic table of personality: Mapping personality scales between the five-factor model and the circumplex model. *Journal of Applied Psychology*, 101, 582–604. <http://dx.doi.org/10.1037/apl0000062>
- Wrzus, C., & Mehl, M. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality*, 29, 250–271. <http://dx.doi.org/10.1002/per.1986>
- Zettler, I., Lang, J. W. B., Hülshager, U. R., & Hilbig, B. E. (2016). Dissociating indifferent, directional, and extreme responding in personality data: Applying the three-process model to self- and observer reports. *Journal of Personality*, 84, 461–472. <http://dx.doi.org/10.1111/jopy.12172>

(Appendices follow)



## Appendix A

### Model Specification for the Trait Variability Tree Models (TVTMs) Used in the Studies

The model specification for the trait variability tree model (TVTM) in Study 1 and Study 2 is as follows:

$$\begin{aligned}
 \Pr(\gamma_{ij} = \text{Strongly disagree}) &= \Phi(\gamma_j^{(I)} + \theta_i^{(I)}) [1 - \Phi(\gamma_j^{(II)} + \theta_i^{(II/IId)})] [1 - \Phi(\gamma_j^{(III)} + \theta_i^{(-I)})] \\
 \Pr(\gamma_{ij} = \text{disagree}) &= \Phi(\gamma_j^{(I)} + \theta_i^{(I)}) [1 - \Phi(\gamma_j^{(II)} + \theta_i^{(II/IId)})] \Phi(\gamma_j^{(III)} + \theta_i^{(-I)}), \\
 \Pr(\gamma_{ij} = \text{neutral}) &= 1 - \Phi(\gamma_j^{(I)} + \theta_i^{(I)}) \\
 \Pr(\gamma_{ij} = \text{agree}) &= \Phi(\gamma_j^{(I)} + \theta_i^{(I)}) \Phi(\gamma_j^{(II)} + \theta_i^{(II/IId)}) \Phi(\gamma_j^{(III)} + \theta_i^{(-I)}) \\
 \Pr(\gamma_{ij} = \text{Strongly agree}) &= \Phi(\gamma_j^{(I)} + \theta_i^{(I)}) \Phi(\gamma_j^{(II)} + \theta_i^{(II/IId)}) [1 - \Phi(\gamma_j^{(III)} + \theta_i^{(-I)})],
 \end{aligned}$$

In this model specification,  $\Phi$  is a short form of the cumulative standard normal distribution. The coding of the pseudoitems is as follows: Pseudoitem I is coded 0 when indifference is endorsed, pseudoitem II is coded 1 when direction is endorsed, and pseudoitem III is coded 0 when intensity is endorsed. The model describes the probability that person  $i$  selects item  $j$ 's midpoint response category "neutral" as a function of the item difficulty parameter  $\gamma_j^{(I)}$  for pseudoitem I and the person's latent variability trait  $\theta_i^{(I)}$ . Pseudoitem I is coded 0 when the "neutral" option is endorsed and 1 when the

"neutral" option is not endorsed. When the "neutral" option is not chosen, the two other pseudoitems determine what happens. Pseudoitem II captures the direction of the response, and pseudoitem III captures its intensity or extremity.  $\theta_j^{(II)}$  is the tendency of the person to endorse the positive or direction of a trait and is functionally similar to the mean-level scores.  $\gamma_j^{(III)}$  captures the item difficulty for the direction of the trait. In Study 1, we used a model with two mean traits so the model includes two tendencies ( $II$ s for sociability and  $IId$  for dutifulness) based on which content the item captures. Finally, intensity responses are a function of the item difficulty for intensity  $\gamma_j^{(III)}$  and the person's latent variability trait  $\theta_i^{(-I)}$ . Pseudoitem III is coded 1 when an "agree" or "disagree" response is chosen and 0 when an "strongly agree" or "strongly disagree" response is endorsed. The latent traits are allowed to correlate with covariance matrix  $\Sigma_1$ . Item-level variability can be modeled as fixed (i.e., persons as random, items as fixed, cf. Böckenholt, 2012) or random effects (e.g., De Boeck, 2008; De Boeck & Partchev, 2012; Zettler et al., 2016) in this type of tree model. For the purpose of simplicity, we report the model for random item-effects in Appendix B.

The model specification in Study 3 slightly differed from the model specification in Study 1 and Study 2. Specifically, Study 3 missed Pseudoitem III and included only one trait mean dimension. Accordingly, these model components are omitted from the model specification. The model estimates for all studies are presented in Appendix B.

(Appendices continues)

## Appendix B

### Model Estimates for the Trait Variability Tree Models (TVTMs) Used

Estimates	Study 1/Study 2	Study 3
Fixed effects		
Pseudoitem I	.92	.43
Pseudoitem II	.31	.85
Pseudoitem III	.53	—
Item random effects		
$\sigma_{\text{Pseudoitem I}}$	.31	.54
$\sigma_{\text{Pseudoitem II}}$	.92	.77
$\sigma_{\text{Pseudoitem III}}$	.38	—
$r_{\text{Pseudoitem I, Pseudoitem II}}$	.22	.02
$r_{\text{Pseudoitem I, Pseudoitem III}}$	−.93	—
$r_{\text{Pseudoitem II, Pseudoitem III}}$	−.15	—
Person random effects		
$\sigma_{\text{Sociability}}$	.21	—
$\sigma_{\text{Dutifulness}}$	.33	—
$\sigma_{\text{Personal initiative}}$	—	.12
$\sigma_{\text{Variability}}$	.39	.49
$r_{\text{Sociability, Dutifulness}}$	.12	—
$r_{\text{Sociability, Variability}}$	.05	—
$r_{\text{Dutifulness, Variability}}$	−.18	—
$r_{\text{Personal initiative, Variability}}$	—	.37
Model fit		
logLikelihood	−29,999.36	−986.20
df	15	8

*Note.*  $n = 58,507$  observations from  $j = 235$  persons and  $i = 96$  ratings for Study 1;  $n = 1,768$  observations from  $j = 77$  persons and  $i = 14$  ratings for Study 3.

Received December 7, 2016  
Revision received October 5, 2017  
Accepted October 6, 2017 ■